

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/786,009

DATE: 09/27/2001

TIME: 12:11:57

Input Set : A:\Inteinm.app

Output Set: N:\CRF3\09272001\I786009.raw

3 <110> APPLICANT: XU, Ming-Qun  
4 <120> EVANS, Thomas C.  
5 <120> TITLE OF INVENTION: Intein Mediated Peptide Ligation  
6 <130> FILE REFERENCE: NEB-150PUS  
7 <140> CURRENT APPLICATION NUMBER: 09/786,009  
C--> 11 <141> CURRENT FILING DATE: 2001-04-17  
13 <150> PRIOR APPLICATION NUMBER: 60/102,413  
14 <151> PRIOR FILING DATE: 1999-09-30  
16 <150> PRIOR APPLICATION NUMBER: PCT/US99/22776  
17 <151> PRIOR FILING DATE: 1999-09-30  
19 <160> NUMBER OF SEQ ID NOS: 9  
21 <170> SOFTWARE: PatentIn Ver. 2.0  
23 <210> SEQ ID NO: 1  
24 <211> LENGTH: 43  
25 <212> TYPE: DNA  
26 <213> ORGANISM: Artificial Sequence  
28 <220> FEATURE:  
29 <223> OTHER INFORMATION: Description of Artificial Sequence: the modified  
30 C-terminal splice junction of the intein from the  
31 gyrA gene of Mycobacterium xenopi  
33 <400> SEQUENCE: 1  
34 gggttgctcag ccagcctact ggcctcaccg gttgatagct gca 43  
36 <210> SEQ ID NO: 2  
37 <211> LENGTH: 39  
38 <212> TYPE: DNA  
39 <213> ORGANISM: Artificial Sequence  
41 <220> FEATURE:  
42 <223> OTHER INFORMATION: Description of Artificial Sequence: the  
43 complementary strand of the C-terminal splice  
44 junction of the modified intein from the gyrA  
45 gene of Mycobacterium xenopi  
47 <400> SEQUENCE: 2  
48 gctatcaacc ggtgaggcca gtagcgttgc tgacgaacc 39  
51 <210> SEQ ID NO: 3  
52 <211> LENGTH: 68  
53 <212> TYPE: DNA  
54 <213> ORGANISM: Artificial Sequence  
56 <220> FEATURE:  
57 <223> OTHER INFORMATION: Description of Artificial Sequence: the polylinker  
58 sequence inserted upstream of the modified intein  
59 from the gyrA gene of Mycobacterium xenopi  
61 <400> SEQUENCE: 3  
62 tcaatctag acatatggcc atagggtggca cccgcctcga gggctcttcc tgcattcagg 60  
63 gacatgca 68  
64 <210> SEQ ID NO: 4  
65 <211> LENGTH: 69  
66 <212> TYPE: DNA

ENTERED

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/786,009

DATE: 09/27/2001

TIME: 12:11:57

Input Set : A:\Inteinm.app

Output Set: N:\CRF3\09272001\I786009.raw

67 &lt;213&gt; ORGANISM: Artificial Sequence

69 &lt;220&gt; FEATURE:

70 &lt;223&gt; OTHER INFORMATION: At position 41, "H" = A or C or T.

72 &lt;220&gt; FEATURE:

73 &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: the

74 complementary strand of the polylinker inserted

75 upstream of the modified intein from the gyrA

76 gene of Mycobacterium xenopi

78 &lt;400&gt; SEQUENCE: 4

79 ctatgtgcata tccgtgatg caggaagagc cctcgaggcg hgcgcgcacc catggccata 60

80 tctctagat 69

82 &lt;210&gt; SEQ ID NO: 5

83 &lt;211&gt; LENGTH: 6509

84 &lt;212&gt; TYPE: DNA

85 &lt;213&gt; ORGANISM: Artificial Sequence

87 &lt;220&gt; FEATURE:

88 &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: pTXB1 plasmid

89 sequence containing the modified intein from the

90 gyrA gene of Mycobacterium xenopi

92 &lt;400&gt; SEQUENCE: 5

93 aactaagctca ggtggcactt ttgggggaaa tgtgcgcgga acccctatctt gttttatctt 60

94 cttaatacat tcaaatatgt atccgctcat gagacaataa cccgtataaa tgcctcaata 120

95 atattgaaaa aggaagaula tgagtattca acatttcctg gtgcgcctta ttcccttttt 180

96 tggggcattt tgccttcctg ttattgctca ccagaaaaag ctgggaaag taaaagatgc 240

97 tgaagatcag ttgggtgcac gagtgggtta cctcgaactg gatctcaaca ggggtaagat 300

98 ccttgagagt ttgcgcgcgc aagaacgttc tccaatgatg agcactttta aagttctgct 360

99 atgtggcgcg gtattatccc gtgttgacgc cgggcaagag caactcggtc gcgcataca 420

100 ctattctcag aatgacttgg ttgagtactc accagtcaca gaaaagcacc ttaaggatgg 480

101 catgacagta agagaattat gcagtgtctc cataaccatg agtgataaca ctgcgggcaca 540

102 cttaactctg acaacgatcg gaggacagaa ggagctaaac gcttttttgc acaacatggg 600

103 ggatcatgta actcgccttg atcgttggga accggagctg aatgaagcca taccaaaaga 660

104 cgaagctgac accacgatgc ctgtagcaat ggcaacaaag ttgcgcacaa tattaactgg 720

105 cgaactaact actctagctt ccgggcacaa attaatagac tggatggagg cggataaagt 780

106 tgcaggacca cttctgcgct cggcccttcc ggttggttgg ttattgtctg ataaatctgg 840

107 agccgggtgag cgtgggtctc ggggtatcat tgcagcactg gggccagatg gtaagccctc 900

108 ccgtatcgta gttatctaca cgaacggggag tcaggcaact atggatgaa gaaatagaca 960

109 gatcgtgag ataggtgctt cactgattaa gcatttgtaa ctgtcagacc aagtttactc 1020

110 aatataactt tagattgatt taccocggtt gataatcaga aaagcccaaa aaacaggaa 1080

111 attgtataag caaatattta aattgtaaaac gttaatattt tgttaaaatt cgcgttaaat 1140

112 ttttgttaaa tcagctcatt tttaaaccaa taggcagaaa tggcaaaat cctttataaa 1200

113 tcaaaaagat agcccaagat aggtttgagt gttgttcag tttggaacaa gagtccacta 1260

114 taaagaacg tggactccaa cctcaaaagg cgaaaaacag tctatcaggy cttatccaca 1320

115 tcaatgaa cttatcccaa atcaagcttt ttgggttoga ggtgcggtta agcactaaat 1380

116 cggaaacctt aagggagccc ccgatttaga gtttjaggy aaaaatcggy gaagcttacc 1440

117 agtttgaa ctttttggc caaadgaagc ggcgttaggy cgttggcaag tctagcggtc 1500

118 accgtgcgcg taaccaccac accgcgcgcg cttaataagg cgttcccaaa cgttcaaaa 1560

119 gatctaggty aagatctttt ttgataatct catgacacaa atcccttaac gtgagttttc 1620

120 gttccactga ggcgcagacc ccgtagaaaa gatcaaaagga tcttcttgag atcctttttt 1680

121 tctgcgcgta atctgtgtgt tgcacacaaa aaaaaccacc ctaccagcgy tggtttgttt 1740

## RAW SEQUENCE LISTING

DATE: 09/27/2001

PATENT APPLICATION: US/09/786,009

TIME: 12:11:57

Input Set : A:\Inteinm.app

Output Set: N:\CRF3\09272001\I786009.raw

```

122 gccggatcaa gagctaccaa ctcttttttc gaaggtaact ggttccagca gagcgcaqat 1800
123 accaaatact gtccttctag tgtagccgta gttaggccac caottcaaga actctgtage 1860
124 accgcctaca tacctcgctc tgctaactct gttaccagtg gctgctgcca gtggcgataa 1920
125 gtctgtgtctt accgggttgg actcaagacg atagttacgg gataaggcgc agcggtcggg 1980
126 ctgaacgggg ggttcgtgca cacagcccag cttggagoga acgacctaca ccgaactgag 2040
127 atacctacag cgtgagctat gagaaagcgc caagcttccc gaaggagaga aggcggacag 2100
128 gtatccggta agcgccaggg tcggaacagg agagcgcaag agggagcttc cagggggaaa 2160
129 cgcctgggat ctttatagtc ctgtcgggtt tcggcacctc tgacttgago gtogattttt 2220
130 gtgatgctcg tcaggggggc ggaacctatg gaaaaacgcc agcaaacggg cctttttacg 2280
131 gttcctggcc ttttgctggc cttttgctca catgtttctt cctgctttat cccctgatto 2340
132 tgtggataac cgtattacgg cctttgagtg agctgatacc gctcgcgcga gcggaacgac 2400
133 cgagcgccgc gaytcagtga gcgaggaagg tatggtgcac tctcagtaaa atctgtcttg 2460
134 atgcgcgata gtttaagccag tatacactcc gctatcgcta cgtgaactgg tcctggctgc 2520
135 gccccgacac ccgccaacac ccgctgacgc gccctgaagg gcttgtctgc tcccggcctc 2580
136 cgtttacaga caagctgtga ccgtctccgg gagctgcctg tgtcagaggt tttcacctgc 2640
137 atcacgaaa ccgcgcaggc agctgcggta aagctcatca gctgggtcgt gcagcgatto 2700
138 acagatgtct gctgtttcat ccgcgtccag ctcgttgagt ttctccagaa gcgttaatgt 2760
139 ctggctttctg ataaagcggg ccattgttaag ggcggttttt tctgttttgg tcaottgatg 2820
140 cctccgtgta agggggaatt tctgttcatg ggggtaatat taccgatgaa acgagagagg 2880
141 atgctcagca taagggttac tgatgatgaa catgcccggg taactggaaag ttgtgagggt 2940
142 aaacaaactg cgttatggat gcggcgggac cagagaaaaa tcactcaggg tcaatgccag 3000
143 ccgaaacgca gcaagacgta gccagccggc tcggccgcga tgcggcgcat aatggcctgc 3060
144 ttctcgccga aacgttttgt gcggggacca gtgacgaagg ctgagcgag ggtgtgcaag 3120
145 ccttcgaata ccgcaagcga caggccgcat atcgctcgcg tcacgcgaaa ggggtctctg 3180
146 ccgaaaaatg accgagagcg tcgggcgacc tgtctcagca gttgcctgat aaagaagaca 3240
147 gtcataaagt cggcgacgat agtcatgcgc ccgcgcgcac ggaaggagct gaactgggtg 3300
148 aaggctctca agggcatcgg tcgagatccc ggtgcctaata gactgagcta acctacatta 3360
149 attgctttgc gctcactgac cgttttccag tcgggaaaac tgtcgtgcca gctgcattaa 3420
150 tgaatcggcc aacgcgcggg gagaggcggg ttgcgtattg ggcgcacagg tggttttctt 3480
151 tttcacccag gagacgggca acagctgatt gcccttcaac gctgggcctt gagagagttg 3540
152 cagcaagcgg tcacagctgg tttgcgccag caggcgaaaa tctgttttga tgggtggttaa 3600
153 cggcgggata taacatygac tgtcttcggg atcgctctat ccactaccc agatatccgc 3660
154 accaacgcgc agcccggaat cggtaatggc gcgcattggc ccagcgcca tctgatcgtt 3720
155 ggcacccagc atcgacgtgg gaacgatgac ctcattoagc atttgcctgg tttgttgaaa 3780
156 accggacatg gcactccagt cccttcccgc ttccgctatc ggttgaattt gattgcaggt 3840
157 gagatattta tgcacgcccag ccagacgcag acgcgcgcag acagaaatta atgggcgcgc 3900
158 taacagcgcg atttgcctgg gacccaatgc gaccagatgc tcacgcgcca gtgcgctacc 3960
159 gttctcatgg gagaaaataa tactgtttgat ggtgtctgg tcagagacat caagaaataa 4020
160 ccgcgggaaca ttagtgcagg cagcttccac agcaatggca tctgtgtcat ccagcgcata 4080
161 gtttaattgat agcccaactg ccgcttgccc gagaagattg tgcacgcgcg ctttacaggg 4140
162 ttgcagcgcg cttgtttcta ccctgcacac caccacgcctg gcaacccagt gatcggcgcg 4200
163 agatttaato gcgcgcacaa tttgcgaagg ccgctgcagg gcacagctgg aatggcaac 4260
164 gccaatcagc aacaaatgtt tgcgcacaa tctttgtgac accgggttgg gaatgtaatt 4320
165 cagctccgac atgcgcgctt ccacttttcc ccggttttcc gcaaaatgtt ggttgcctta 4380
166 gtttaacaa ggggaaatgt tctgataada cacaacggca taotctgaga ctctgtataa 4440
167 cgttaactggt ttcacattca ccacccctgaa ttgactctct tccggcctat atcttccat 4500
168 acccgcaaa gttttggggt attcgatggt ctcccgcatc tcgaagctct cctttatggg 4560
169 actcctgcat taggaagcag ccagtagta ggttgaggcc gttgagcaca gcgcgcgcaa 4620
170 ggaatggtgc atgcgcgctt ttgctcttca aqaattaatt cccaattcca ggcatacaat 4680

```

## RAW SEQUENCE LISTING

DATE: 09/27/2001

PATENT APPLICATION: US/09/786,009

TIME: 12:11:57

Input Set : A:\Inteinm.app

Output Set: N:\CRF3\09272001\I786009.raw

```

171 aaaaagaaag gctcagtcga aagactgggc ctttcgiltt atctgttggt tctcggtgaa 4740
172 cgcctctctg agtaggacaa atccgcggg agcggatttg aacgttgoga agcaacggcc 4800
173 cggaggggtg cgggcaggac gccgcgcata aactgcaggg aattaattcc aggcacaaaa 4860
174 taaaacgaaa ggctcagtcg aaagactggg cctttcgitt tatctgttgt ttgtcggtga 4920
175 acgctctctt gactaggaca aatccgcggg gacgcgattt gaaagtttgc aagcaacggc 4980
176 cgggaggggt gcgggcagga cgcgcgcgat aaactgcag gaattaattc caggcatcaa 5040
177 ataaaacgaa aggctcagtc gaaagactgg gcctttcggt ttatctgttg ttgtcggtg 5100
178 aacgctctcc tgactaggac aaatccgcgg ggagcggatt tgaacgttgc gaagcaacgg 5160
179 cccgcgagggt gcgggcagg acgcgcgcga taaactgcga ggaattaatt ccaggcatca 5220
180 aataaaacga aaggctcagt cgaagactg ggcccttcgt ttatctgtt gttgtcggt 5280
181 gaacgctctc ctgactagga caaatccgac gggagcggat ttgaacgttg cgaagcaacg 5340
182 gccgcgaggg tggcgggcag gacgcgcgac ataaactgac aggaattggg gatcggaatt 5400
183 aattccgggt ttaaacggg gatctcgat ccgcgaaatt aatacgaact actatagggg 5460
184 aattgtgagc ggataacaat tccctctag aaataatttt gtttaacttt aagaaggaga 5520
185 tatacatatg gctagctcgc gactcgacgg cggcgcgcaa ttcctcgagg gctcttctg 5580
186 catcacggga gatgcaatg ttgcctaac cgagggcgag tcggtacgca tcgcgcacat 5640
187 cgtgcggggt gcgcgcgcca acagtgcaca cgcctatgac ctgaaagtcc ttgacgggca 5700
188 tggcaatccc gtgctcgccg accggtgtgt ccactccggc gacatccgg tgtaacgggt 5760
189 gcgtacggtc gaaggtctgc gtgtgacggg caccgcgaa caccggttgt tgtgtttggt 5820
190 cgaactcgcc ggggtgcgca cctgctgtg gaagctgac gacgaaatca agcggggcga 5880
191 ttaacgggtg atcaaacgca gcgcattcag cgtcgactgt gcaggttttg cccgcgggaa 5940
192 aaccgaattt gcgcgcacaa cctacacagt cggcgtccct ggaactggtg gtttcttgga 6000
193 agcaaacgac cgaagacgca acgcgcgaag tatcgccgac gactgacgc acgggcgggt 6060
194 ctactacggc aaagtgcgca gtgtcacgca cgcggcggtg cagcgggtgt atagccttgc 6120
195 tctcgacacg gcagaccacg cgtttatcac gaacgggttc gtcagcagc ctactggcct 6180
196 caccggtctg aactcaggcc tcacgacaaa tccgtgtgta tccgcttggc aggtcaacac 6240
197 agcttatact gcgggacaat tggtcacata taacggcaag acgtataaat gtttgacgac 6300
198 ccacacctcc ttggcaggat gggaaaccac caacgttccg gccttgtggc agcttcaatg 6360
199 actgcaggaa ggggatccgg ctgctaacaa agcccgaaag gaagctgagt tggctgtgc 6420
200 caccgctgag caataactag cataaccctc tggggcctct aaaacgggtct tgaggggttt 6480
201 ttgtgtgaaa ggaggaaacta tatccggat 6509

```

203 &lt;210&gt; SEQ ID NO: 5

204 &lt;211&gt; LENGTH: 30

205 &lt;212&gt; TYPE: PRT

206 &lt;213&gt; ORGANISM: Artificial Sequence

208 &lt;220&gt; FEATURE:

209 &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: synthetic

210 peptide

212 &lt;400&gt; SEQUENCE: 6

213 Dns Ala Tyr Lys Thr Thr Gln Ala Asn Lys His Ile Ile Val Ala Cys

214 1 5 10 15

215 Gln Gly Asn Pro Tyr Val Pro Val His Phe Asp Ala Ser Val

216 20 25 30

217 &lt;210&gt; SEQ ID NO: 7

218 &lt;211&gt; LENGTH: 18

219 &lt;212&gt; TYPE: PRT

220 &lt;213&gt; ORGANISM: Artificial Sequence

222 &lt;220&gt; FEATURE:

223 &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: the amino acid

## RAW SEQUENCE LISTING

DATE: 09/27/2001

PATENT APPLICATION: US/09/786,009

TIME: 12:11:57

Input Set : A:\Inteinm.app

Output Set: N:\CRF3\09272001\I786009.raw

```

227     sequence deduced from the polylinker region of
228     pTXB1
230 <400> SEQUENCE: 7
231 Met Ala Met Gly Gly Arg Leu Glu Gly Ser Ser Cys
232   1             5             10
235 <210> SEQ ID NO: 8
236 <211> LENGTH: 42
237 <212> TYPE: DNA
238 <213> ORGANISM: Artificial Sequence
240 <220> FEATURE:
241 <223> OTHER INFORMATION: Description of Artificial Sequence: polylinker ✓
242     region upstream of the modified intein from the
243     gyrA gene of Mycobacterium xenopi in pTXB1
245 <400> SEQUENCE: 8
246 catatggcca tgggtggcgg ccgcctcgag ggctcttctt gc           42
248 <210> SEQ ID NO: 9
249 <211> LENGTH: 7
250 <212> TYPE: PRT
251 <213> ORGANISM: Artificial Sequence
253 <220> FEATURE:
254 <223> OTHER INFORMATION: Description of Artificial Sequence: synthetic ✓
255     peptide
257 <400> SEQUENCE: 9
258 Cys Asp Pro Glu Lys Asp Ser
259   1             5

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/786,009

DATE: 09/27/2001

TIME: 12:11:58

Input Set : A:\Inteinm.app

Output Set: N:\CRF3\09272001\I786009.raw

L:11 M:271 C: Current Filing Date differs, Replaced Current Filing Date